2016 GUIDELINES REGARDING THE USE OF TECHNOLOGY-ASSISTED REVIEW

The Coalition of Technology Resources for Lawyers (CTRL) is pleased to publish the *2016 Guidelines regarding the Use of Technology-Assisted Review*. The *2016 Guidelines* build on the inaugural *Guidelines* published in 2014, which considered some of the divisive issues surrounding the use of Technology-Assisted Review ("TAR"). Recently cited by some of the leading authorities on civil discovery issues, the *Guidelines* focused on aspects of a defensible TAR workflow. They also considered whether and to what extent counsel should disclose its use of TAR, along with the benefits and drawbacks of entering into a stipulated TAR use protocol.

The *2016 Guidelines* expand on these issues. They examine whether and to what extent a defensible workflow may include other methodologies used in combination with TAR. They also explore some of the principal methods for developing seed sets and training a TAR algorithm. They next consider the issue of disclosure in the context of both traditional litigation and government investigations. Finally, they analyze new developments in the jurisprudence on stipulated use protocols.

The *2016 Guidelines* spotlight these issues since—at some level—they must be addressed when considering the use of TAR. CTRL believes the *2016 Guidelines* will ultimately help attorneys, clients, and judges gain a better understanding of the issues surrounding the use of TAR.

CTRL wishes to thank all those who have contributed to the development of the *2016 Guidelines*. Should you wish to contribute to the ongoing process of refining, updating, and improving the *Guidelines*, please visit our website at http://www.ctrlinitiative.com/home/protocol/ to submit comments, suggestions, or revisions. If you would like to get more directly involved with the *Guidelines* or other CTRL projects, please email us at phil@ctrlinitiative.com.

Philip J. Favro
Director of Legal Education and Resources
Coalition of Technology Resources for Lawyers
August 2016

INTRODUCTION

Like any new and disruptive technology, the use of TAR[1] has been both contagious and controversial since its introduction to the discovery process.[2] On the one hand, TAR has found welcome recipients in clients, counsel, and the courts, all of whom are seeking to expedite the ESI search and review process.[3] Lawyers and litigants have additionally gravitated toward TAR given its utility in identifying the key documents required to establish their claims or defenses.[4]

Nevertheless, there continue to be disagreements regarding what is TAR,[5] when it should be used,[6] and the process for how to successfully implement it into a discovery workflow.[7] Moreover, the judicial opinions that address TAR are based on specific fact patterns that make general application for practitioners difficult.[8]

In an effort to dispel confusion over these issues and to help provide informed direction on the use of TAR, CTRL has prepared a few guidelines that should aid counsel, clients, and the courts through the decision-making process on the following key issues:

- What are the touchstones of a defensible TAR use plan?
- What are some essential aspects of a TAR workflow?

---

[1] In these Guidelines, TAR shall mean and refer to a process for selecting and ranking a collection of documents using a computerized system that incorporates the decisions that lawyers have made on a smaller set of documents and then applies those decisions to the remaining universe of documents. Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review, with a Foreword by John M. Facciola, U.S. Magistrate Judge*, 7 FED. CTS. L. REV. 1, 38 (2013). *See also* Bennett Borden & Jason R. Baron, *Finding the Signal in the Noise: Information Governance, Analytics, and the Future of Legal Practice,* 20 RICH. J.L. & TECH. 7, ¶ 10 (2014). While TAR is often used interchangeably with terms such as "predictive coding" and "computer-assisted review," TAR "now seems to be the preferred term of art" and is therefore used throughout *2016 Guidelines* in lieu of the term "predictive coding." Rio Tinto PLC v. Vale S.A. (*Rio Tinto I*), 306 F.R.D. 125, 126 (S.D.N.Y. 2015).

[2] *See, e.g.,* In re Domestic Drywall Antitrust Litig., 88 Fed. R. Serv. 3d 966 (E.D. Pa. May 12, 2014) (touting the benefits of TAR as a new and "sophisticated" discovery search methodology); Dana Remus, *The Uncertain Promise of Predictive Coding,* 99 IOWA. L. REV. 1691, 1695 (2014) (suggesting that lawyers "proceed with deliberate care in the use and adoption of predictive-coding technologies.").

[3] Nat'l Day Laborer Org. Network v. U.S. Immigration & Customs Enforcement Agency (*NDLON*), 877 F. Supp. 2d 87, 109 (S.D.N.Y. 2012) ("[P]arties can (and frequently should) rely on . . . machine learning tools to find responsive documents."); Gareth Evans & David Grant, *Tools Let Attorneys Follow the Breadcrumbs: Analytics applications can help litigators identify surface patterns that point to the key evidence*, THE NAT'L LAW JOURNAL (Sept. 1, 2014), *available at http://www.gibsondunn.com/publications/Documents/Evans-ToolsLetAttorneysFollowTheBreadcrumbs.pdf*; (describing the capabilities of advanced analytics in discovery); Hon. Patrick J. Walsh, *Rethinking Civil Litigation in Federal District Court*, 40 LITIG. 6, 7 (2013) (urging lawyers to use "21st-century computer technology" including TAR to address digital age discovery issues and stop relying on legacy discovery technologies and methods).

[4] Walsh, *supra* note 3, at 7 ("Their goal should be to have the computer sift through the millions of documents and distill and organize the hundreds or thousands of documents that are critical to the case . . . ."); Charles Yablon & Nick Landsman-Roos, *Predictive Coding: Emerging Questions and Concerns*, 64 S.C. L. REV. 633, 644 (2013) (describing seed set development and its impact on the need "to identify those documents that are most relevant").

[5] Remus, *supra* note 2, at 1706-07 (observing disapprovingly that "the litigation community is uncritically embracing TAR as if its definition is unitary and clear, its accuracy and efficacy well-established.").

[6] Borden, *supra* note 1, at ¶ 17 ("[W]e bow to the reality that in a large class of cases the use of predictive coding is currently infeasible or unwarranted, especially as a matter of cost.").

[7] Maura R. Grossman & Gordon V. Cormack, *Comments on "The Implications of Rule 26(g) on the Use of Technology-Assisted Review,"* 7 FED. CTS. L. REV. 285 (2014) (discussing various disagreements between the authors and other commentators on the proper use of and objectives surrounding TAR).

[8] *Compare* Progressive Cas. Ins. Co. v. Delaney*, No. 2:11-cv-00678, 2014 WL 3563467, at * 10-12 (D. Nev. July 18, 2014) (interpreting case management order provision to prevent the use of TAR) *with* Bridgestone Americas, Inc. v. Int'l Bus. Mach. Corp. (*Bridgestone*), No. 3:13-cv-1196, 2014 U.S. Dist. LEXIS 142525 (M.D. Tenn. July 22, 2014) (allowing the use of TAR despite an arguably contrary case management order).

- Should the use of TAR be disclosed to litigation adversaries?
- What are the benefits and drawbacks of entering into a stipulated TAR use protocol?

Finally, a stipulation and order regarding the use of TAR has also been published on the CTRL website that can be used as a model if required by a court or if the circumstances in a particular matter lend themselves to reaching a stipulation with a litigation adversary.[9] The model stipulation and order is provided with the caveat that pursuing such a course is not necessarily a best practice. Whether a party should enter into a stipulated protocol will depend upon the circumstances of each case, requiring the application of legal judgment.[10]

I. WHAT ARE THE TOUCHSTONES OF A DEFENSIBLE TAR USE PLAN?

Counsel must be prepared to defend its use of TAR just as it would with any other discovery search methodology.[11] Indeed, the TAR process is not subject to any heightened scrutiny above and beyond conventional search methodologies.[12] This means that counsel's use of TAR must accord with the standards of relevance, proportionality, and reasonableness, the traditional touchstones of the discovery process.[13] Indeed, perfection is not the standard, which the extant TAR jurisprudence repeatedly makes clear.[14]

Whether a production satisfies the standards of relevance, proportionality, and reasonableness will ultimately depend on the quality and nature of the responsive information disclosed to an adversary.[15] This means that production results should trump questions of process, even if all responsive information has not been turned over in discovery.[16]

---

[9] *Model Stipulation and Order re Use of Predictive Coding*, COALITION OF TECHNOLOGY RESOURCES FOR LAWYERS (July 22, 2014), *available at* http://www.ctrlinitiative.com/wp-content/uploads/2014/08/Stipulation-and-Order-re-use-of-Predictive-Coding-in-Discovery-FINAL.doc.

[10] *See infra* Part IV; Remus, *supra* note 2, at 1716-19 (cautioning lawyers not to yield the interests of their clients in the quest for reasonable cooperation in the discovery process).

[11] *Bridgestone* (successfully defending the combined use of keywords and TAR). *See also* William A. Gross Const. Associates, Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 134 (S.D.N.Y. 2009) ("This Opinion should serve as a wake-up call to the Bar in this District about the need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or 'keywords' to be used to produce emails or other electronically stored information").

[12] *See Rio Tinto I*, at 129 ("One point must be stressed - it is inappropriate to hold TAR to a higher standard than keywords or manual review. Doing so discourages parties from using TAR for fear of spending more in motion practice than the savings from using TAR for review.").

[13] *See* FED. R. CIV. P. 26(b)(1); 26(b)(2)(3); 26(g)(1); Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *3 (S.D.N.Y. Aug. 1, 2016) ("the standard is not perfection . . . but whether the search results are reasonable and proportional."); Victor Stanley, Inc. v. Creative Pipe, Inc., 269 F.R.D. 497, 523 (D. Md. 2010) (observing that "all permissible discovery must be measured against the yardstick of proportionality."). The recent amendments to Federal Rule of Civil Procedure 26(b)(1) clarify that all discovery must be viewed through the lenses of relevance and proportionality. *See* FED. R. CIV. P. 26 advisory committee's note, 2015 Amendment to subdivision (b)(1) ("Proportional discovery relevant to any party's claim or defense suffices.").

[14] *Hyles*, 2016 WL 4077114 at *3; Dynamo Holdings Ltd. P'Ship v. Comm'r of Internal Revenue (*Dynamo Holdings II*), No. 2685-11 (T. C. July 13, 2016) ("The second myth is the myth of a perfect response. The Commissioner is seeking a perfect response to his discovery request, but our Rules do not require a perfect response. Instead, the . . . Rules require that the responding party make a 'reasonable inquiry' before submitting the response.") Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc., 11-cv-6189, 2014 WL 584300, *2 (S.D.N.Y. Feb. 14, 2014) ("[While] [p]arties in litigation are required to be diligent and to act in good faith in producing documents in discovery . . . . no one could or should expect perfection from this process. All that can be legitimately expected is a good faith, diligent commitment to produce all responsive documents uncovered when following the protocols to which the parties have agreed, or which a court has ordered."); *Da Silva Moore*, 287 F.R.D. at 191 ("While this Court recognizes that computer-assisted review is not perfect, the Federal Rules of Civil Procedure do not require perfection.").

[15] *See Fed. Hous. Fin. Agency*, 2014 WL 584300 at *2; *In re* Biomet M2a Magnum Hip Implant Products Liability Litig. (*Biomet II*), No. 3:12-MD-2391, 2013 WL 6405156, at *1-2 (N.D. Ind. Aug. 21, 2013); *In re* Biomet M2a Magnum Hip Implant Products Liability Litig. (*Biomet I*), No. 3:12-MD-2391, 2013 WL 1729682, at *3 (N.D. Ind. Apr. 18, 2013); Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 191 (S.D.N.Y. 2012).

[16] *See Dynamo Holdings II,* at *8 ("[W]hen the responding party is signing the response to a discovery demand, he is not certifying that he turned over everything, he is certifying that he made a reasonable inquiry and to the best of his knowledge, his response is complete."); Larsen v.

The inquiry for a court is not whether the process was completely transparent, whether the parties were cooperating, or whether the process ensured that all responsive information would be produced.[17] Instead, the proper focus for a court is whether the responding party's discovery efforts were reasonable and proportional under the circumstances given what information was produced in discovery and what is at stake in the litigation.[18]

To be sure, a responding party must be prepared to establish the defensibility of its process. Nevertheless, a court should not immediately proceed to addressing questions of process unless the requesting party can establish legitimate production shortcomings with a fact-specific showing of good cause.[19] Considerations of relevance, proportionality, and reasonableness weigh against inquiring into the responding party's TAR process without good cause.[20]

A corollary to this principle is that a court generally should not mandate or direct the parties to use a particular process or methodology for conducting discovery. Such a position runs counter to the well-established rule that the responding party is best situated to determine how it should search for, review, and produce its responsive documents.[21] This is confirmed by *Hyles v. New York City* and the first opinion from *Dynamo Holdings v. Commissioner of Internal Revenue*, both of which observed that courts are "not normally in the business of dictating to parties the process that they should use when responding to discovery."[22]

Some courts, however, have deviated from or disregarded these principles entirely by unilaterally ordering parties to disclose their TAR process or directing them to enter into a stipulated TAR use protocol.[23] Their purpose in doing so

---

Coldwell Banker Real Est. Corp., No. SAVC 10-00401-AG (MLGx), 2012 WL 359466 at *8 (C.D. Cal. Feb. 2, 2012) ("To require Defendants to repeat this labor merely because Plaintiffs have identified a few alleged discrepancies in the ESI production is simply unreasonable.").

[17] *Dynamo Holdings II,* at *8; *Biomet I*, at *2 ("What Biomet has done complies fully with the requirements of Federal Rules of Civil Procedure 26(b) and 34(b)(2). I don't see anything inconsistent with the Seventh Circuit Principles Relating to the Discovery of Electronically Stored Information. Principle 1.02 requires cooperation, but I don't read it as requiring counsel from both sides to sit in adjoining seats while rummaging through millions of files that haven't been reviewed for confidentiality or privilege.").

[18] *Hyles*, 2016 WL 4077114 at *3; *Biomet I*, at *2-3. *See also* FED. R. CIV. P. 26 advisory committee's note, 2015 Amendment to subdivision (b)(1) ("Proportional discovery relevant to any party's claim or defense suffices."); THE SEDONA CONFERENCE COMMENTARY ON PROPORTIONALITY IN ELECTRONIC DISCOVERY, THE SEDONA CONFERENCE (Jan. 2013), *available at* https://thesedonaconference.org/publication/The%20Sedona%20Conference%20Commentary%20on%20Proportionality.

[19] *See Hyles*, 2016 WL 4077114 at *3; FED. R. CIV. P. 37(a); Abreu v. State of New Jersey, No. 14-cv-716, 2015 WL 9480021, *4 (D.N.J. Dec. 29, 2015) (discussing the general rule that a movant must demonstrate good cause before a court will grant a motion to compel).

[20] *See generally* FED. R. CIV. P. 26 advisory committee's note, 2015 Amendment to subdivision (b)(1).

[21] *Hyles*, 2016 WL 4077114 at *3 ("Under Sedona Principle 6, the City as the responding party is best situated to decide how to search for and produce ESI responsive to Hyles' document requests. . . . it is not up to the Court, or the requesting party (Hyles), to force the City as the responding party to use TAR when it prefers to use keyword searching."); Kleen Products LLC v. Packaging Corp. of Am., No. 10-cv-5711, 2012 WL 4498465, *5 (N.D. Ill. Sept. 28, 2012) ("[T]he Court observed that under Sedona Principle 6, '[r]esponding parties are best situated to evaluate the procedures, methodologies, and techniques appropriate for preserving and producing their own electronically stored information.'"); Ford Motor Co. v. Edgewood Props., 257 F.R.D. 418, 427 (D.N.J. 2009) ("The Sedona Principles wisely state that it is, in fact, the producing party who is the best position to determine the method by which they will collect documents. The producing party responding to a document request has the best knowledge as to how documents have been preserved and maintained."); THE SEDONA CONFERENCE, THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION (*SEDONA PRINCIPLES*) 38 (Jonathan M. Redgrave et al. eds., 2d ed. 2007) (providing under Principle Six that the responding party is best situated to determine how it should search for, review, and produce its responsive documents), *available at* https://thesedonaconference.org/publication/The%20Sedona%20Principles.

[22] Dynamo Holdings Ltd. P'Ship v. Comm'r of Internal Revenue (*Dynamo Holdings I*), 143 T.C. No. 9, 2014 WL 4636526, at *188 (T. C. Sept. 17, 2014); *accord Hyles*, 2016 WL 4077114 at *3 (relying on *Dynamo Holdings I* for the quoted proposition).

[23] *See*, *e.g.*, Boardley v. Household Finance Corporation III (*Boardley*), No. 12-cv-03009-PWG, 2015 U.S. Dist. LEXIS 70098 at *11-12 (D. Md. June 1, 2015) ("Parties requesting ESI discovery and parties responding to such requests are expected to cooperate in the development of search methodology and criteria to achieve proportionality in ESI discovery, including appropriate use of computer-assisted search methodology, such as Technology Assisted Review."); *Bridgestone*, at *2 ("openness and transparency in what Plaintiff is doing will be of critical importance.

is to ensure a more transparent and cooperative approach to discovery, thereby streamlining the process and eliminating questionable search methods that could limit the production of responsive information.[24] While well intentioned, this approach to the discovery process may not be appropriate in many circumstances. [25] It can be particularly problematic in those instances where there are disputes over the combined use of TAR with other search methodologies.[26]

The Use of TAR Together with other Search Methodologies

Since its inception in discovery, lawyers have generally used TAR together with other search methodologies.[27] This is because the circumstances of a particular matter often lend themselves to using other methods with TAR for a document production. Indeed, the nature of the data, text extraction or optical character recognition, images, and numerically based documents may favor a client's use of TAR with other search methodologies in a particular workflow.[28] These issues and others demonstrate that TAR is not necessarily a stand-alone solution and can be used in conjunction with search terms, manual review, or other search methodologies[29] to achieve productions that satisfy standards of relevance, proportionality, and reasonableness.[30]

Nevertheless, requesting parties have sometimes bristled at the use of search terms with TAR. This is particularly the case where a responding party has used "search terms and other criteria . . . to reduce the volume of the Document Universe."[31] Responding parties often use keyword searches to eliminate marginally responsive or non-responsive information given the costs of processing an entire document population with TAR technology.[32] However, because over-inclusive keyword searches could potentially eliminate materially responsive data, requesting parties have urged courts either to forbid responding parties from using keywords with TAR or to require

---

Plaintiff has advised that they will provide the seed documents they are initially using to set up predictive coding. The Magistrate Judge expects full openness in this matter.").

[24] *See* Progressive Cas. Ins. Co. v. Delaney*, No. 2:11-cv-00678, 2014 WL 3563467, at *10-12 (D. Nev. July 18, 2014) ("Progressive is unwilling to engage in the type of cooperation and transparency that its own e-discovery consultant has so comprehensibly and persuasively explained is needed for a predictive coding protocol to be accepted by the court or opposing counsel as a reasonable method to search for and produce responsive ESI.").

[25] *See infra* Parts III, IV.

[26] *See infra* Part I (The Use of TAR Together with other Search Methodologies).

[27] *See generally Biomet I* (approving the defendant's combined use of search terms and TAR where the search terms were first used to remove nonresponsive information, thereby "reducing the universe of documents and attachments from 19.5 million documents to 3.9 million documents").

[28] *See* Elle Byram, *The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery*, 29 SANTA CLARA COMPUTER & HIGH TECH. L.J. 675, 694 (describing several issues and other variables affecting the choice of a particular discovery search methodology).

[29] *See*, *e.g.*, *NDLON,* at 109 ("And beyond the use of keyword search, parties can (and frequently should) rely on latent semantic indexing, statistical probability models, and machine learning tools to find responsive documents."); *Biomet I* (approving the combined use of keyword search terms and TAR); *Bridgestone* (approving the combined use of keyword search terms and TAR).

[30] *See Biomet I* (holding that the defendant's combined use of keyword and TAR search methodologies satisfied its response obligations under Rule 26 and Rule 34).

[31] *Rio Tinto I*, at 132 (providing in the parties' stipulated TAR use protocol that the responding party may use search terms to decrease the number of potentially responsive documents if it determines such an approach "to be reasonable and appropriate").

[32] *See Biomet I. See also* Yablon, *supra* note 4, at 677 ("The question has arisen whether a responding party may unilaterally create and deploy keyword search terms to winnow down a pool of data, or whether there must be some form of agreement. . . . a good faith, unilateral approach to the development of keywords for culling electronic documents has been considered defensible").

complete transparency in the use of search terms.[33] At least one court has apparently agreed with this position and prohibited a responding party from using keywords to decrease the universe of potentially responsive information.[34]

The majority rule on this issue nonetheless remains that litigants can use search terms to remove non-responsive materials prior to running TAR on the remaining document population.[35] This generally accords with notions of proportionality and reasonableness due to the high costs often associated with the ingestion of data into the TAR workflow. It is also consistent with the generally accepted discovery principle that responding parties are best situated to determine how they should search for, review, and produce their responsive documents.[36] Thus, courts will generally approve the combined use of TAR with other methodologies designed to reasonably reduce the size of the document population so long as the process results in productions of highly relevant information that are proportional under the circumstances.[37] If these standards are not satisfied, responding parties should be prepared to re-do their productions despite the potential cost burdens of doing so.[38]

II. WHAT ARE SOME ESSENTIAL ASPECTS OF A TAR WORKFLOW?

There are various issues that must be considered in connection with a TAR workflow.[39] To be sure, those issues will inevitably vary depending on the quality and nature of the technology adopted by the responding party, not to mention the facts and circumstances of a particular matter. For example, the time when counsel decides to apply TAR or other search methodologies to the universe of potentially responsive information – particularly when the search and review process begins before the Rule 26(f) conference – may impact workflow defensibility. This is

---

[33] *See NDLON,* at 109 ("There is increasingly strong evidence that '[k]eyword search[ing] is not nearly as effective at identifying relevant information as many lawyers would like to believe.'").

[34] *See, e.g.*, Order Re: Implementation of Predictive Coding Regimen, Indep. Living Ctr. of S. Cal. v. City of L.A., 2:12-cv-00551-FMO-PJW, at *1 (C.D. Cal. June 13, 2014) ECF No. 371 (forbidding the defendant from using a particular search methodology in conjunction with TAR).

[35] *Dynamo Holdings I*, at *194 ("Mr. Scarazzo concluded that petitioners' approach would reduce the universe of information on the tapes using criteria set by the parties to minimize review time and expense and ultimately result in a focused set of information germane to the matter."); *Biomet I*, at *2. *But see* Rio Tinto PLC v. Vale S.A. (*Rio Tinto II*), 14-cv-03042-RMB-AJP, 2015 WL 4367250, *2 (S.D.N.Y. July 15, 2015) ("The Court itself felt bound by the parties' protocol, such as to allow keyword culling before running TAR, even though such pre-culling should not occur in a perfect world.").

[36] Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *3 (S.D.N.Y. Aug. 1, 2016); *Rio Tinto I*, at 126, n. 2 (citing and discussing Sedona Principle Six); Kleen Products LLC v. Packaging Corp. of Am., No. 10-cv-5711, 2012 WL 4498465, *5 (N.D. Ill. Sept. 28, 2012); Ford Motor Co. v. Edgewood Props., 257 F.R.D. 418, 427 (D.N.J. 2009); *SEDONA PRINCIPLES*, at 38.

[37] *See Biomet I*, at *3; Moore v. Publicis Groupe, 287 F.R.D. 182, 193 (S.D.N.Y. 2012) ("As with keywords or any other technological solution to ediscovery, counsel must design an appropriate process, including use of available technology, with appropriate quality control testing, to review and produce relevant ESI while adhering to Rule 1 and Rule 26(b)(2)(C) proportionality."); Hon. Craig B. Shaffer, *"Defensible" by What Standard?,* 13 SEDONA CONF. J. 217, 218 (2012) ("Ultimately, a technology-assisted review process must comport with the requirements of the Federal Rules of Civil Procedure, be proportionate to the claims, defenses and circumstances of the particular case, and be reasonably transparent to the court and opposing parties."). This notion is also supported in jurisdictions beyond the United States such as England and Wales. *See* Pyrrho Investments v MWB Property, 2016 EWHC 256 (Ch) (Feb. 16, 2016) ("In my judgment the estimated costs of using the [TAR] software are proportionate."); Brown v BCA Trading, [2016] EWHC 1464 (CH) (May 17, 2016) (approving the use of TAR and observing that its use is consistent with proportionality standards).

[38] *Hyles*, 2016 WL 4077114, *3 ("If Hyles later demonstrates deficiencies in the City's production, the City may have to re-do its search."). *See also* Order Re: Implementation of Predictive Coding Regimen, Indep. Living Ctr. of S. Cal. v. City of L.A., 2:12-cv-00551-FMO-PJW, at *1 (C.D. Cal. June 13, 2014) ECF No. 371.

[39] *See* Nicholas Barry, Note, *Man Versus Machine Review: The Showdown between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology*, 15 VAND. J. ENT. & TECH. L. 343, 354-55 (2013) (describing a TAR workflow and related issues associated with its implementation).

because decisions on search methodologies and discovery technologies are frequently made before a complaint is filed or opposing counsel is even identified.[40]

Irrespective of these and other issues, however, the ultimate defensibility of a TAR workflow is generally not dependent on the so-called "best practices" advanced by any particular eDiscovery technology provider.[41] Instead, the defensibility of such a workflow should be evaluated on whether the production satisfies the touchstones of relevance, proportionality, and reasonableness.[42] Three general areas that are fundamental to doing so are discussed below.

Prevalence

First, counsel should confirm that it has accurately determined the prevalence of responsive information within the universe of documents.[43] As an initial step in this process, counsel should be free to consider other search methodologies in a reasonable and proportional manner to help narrow the subset of potentially responsive documents.[44] After doing so, counsel should then ensure that the "control set" or "sample set" of documents reflects the approximate percentage of responsive data found within that universe of documents.[45] Having an accurate reading of prevalence is essential to establishing overall search and production objectives for the TAR process.[46] If the prevalence evaluation is off the mark, the ultimate evaluation of the review and production quality will be difficult.[47] The production could be under-inclusive, leaving potentially key, responsive information out of the production.[48] Alternatively, the production could be over-inclusive, resulting in the production of too much marginally responsive or non-responsive data.[49]

Seed Sets and Training

Next, counsel should prepare a training or seed set of documents designed to elicit responsive information from the universe of documents.[50] This step is perhaps the most crucial aspect in establishing a properly functioning and

---

[40] Clients who have made a significant investment in in-house tools should be able to leverage that investment to determine how to produce responsive information in a manner that reasonably satisfies their discovery obligations without having to seek permission to do so from litigation adversaries. *See, e.g.*, *Bridgestone*, at *1-2 (holding that the responding party was free to combine TAR and keyword search methodologies to accomplish its production of documents despite arguably contrary language from a case management order); SEDONA PRINCIPLES, at 38.

[41] *But see* Progressive Cas. Ins. Co. v. Delaney, No. 2:11-cv-00678, 2014 WL 3563467, at *4, *8-11 (D. Nev. July 18, 2014) (denying the plaintiff's request to use TAR to assist with its production of documents since (among other reasons) its proposed TAR methodology did "not comply with all of [its technology vendor's] recommended best practices.").

[42] *Bridgestone,* at 2 ("In the final analysis, the uses of predictive coding [are] a judgment call, hopefully keeping in mind the exhortation of Rule 26 that discovery be tailored by the court to be as efficient and cost-effective as possible.").

[43] Prevalence is synonymous with terms such as richness and yield. *See* Grossman, *supra* note 1, at 26.

[44] *See supra* Part I. *Compare* Yablon, *supra* note 4, at 638-39 (describing the cost benefits of removing clearly non-responsive documents from the universe of potentially responsive information), *with Rio Tinto II* (observing in dicta that the use of search terms for "pre-culling should not occur in a perfect world").

[45] *See* Yablon, *supra* note 4, at 640.

[46] *Id*.

[47] *Id*.

[48] *Id*.

[49] *Id*.

[50] *See* Daniel Martin Katz, *Quantitative Legal Prediction - Or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 946 (2013) (explaining that TAR "approaches are inductive and typically involve the seeding

defensible TAR workflow.[51] Indeed, the effectiveness of a TAR process and its ability to satisfy the traditional discovery touchstones of relevance, proportionality, and reasonableness[52] will likely turn on what documents counsel uses to train the algorithm.[53]

A seed set is a proportionately small subset of data that contains examples of the categories of information being sought.[54] The TAR algorithms use the characteristics of the seed set to find similar documents.[55] Those seed documents are then run through the TAR process to train the algorithm to identify documents for production.[56]

Judgmental Sampling

There are two general approaches for developing a seed set.[57] The first is typically referred to as judgmental sampling.[58] Judgmental sampling involves customizing a seed set with specific documents or classes of information that, when submitted into a TAR workflow, are targeted to uncover additional documents whose content is similar to that of the seed materials.[59] Such a seed set could be comprised of highly relevant information, marginally responsive materials, privileged communications, non-responsive data, or a combination of these or even other items.[60] The specific compilation of a seed set, the number of documents in a seed set, and the number of seed sets created will depend on the facts and circumstances of a particular matter.[61]

---

of the algorithm with training (or labeled) data from which the machine infers the 'true' function for assigning a document to a particular group (i.e., relevant versus not relevant).").

[51] *See, e.g.,* John M. Facciola & Philip J. Favro, *Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled To Work Product Protection,* 8 FED. CTS. L. REV. 1, 6 (2015) (discussing generally the importance of seed sets in the TAR process).

[52] *Bridgestone,* at *2-3.

[53] Yablon, *supra* note 4, at 638-39, 643-44 (detailing the importance of the seed set to establishing a reasonable TAR workflow).

[54] Grossman, *supra* note 1, at 29.

[55] Yablon, *supra* note 4, at 638-39, 643-44 ("Judgmental sampling, on the other hand, requires that attorneys with knowledge of the case select documents—already uncovered through discovery—as 'seeds' that they have determined are clearly fitting or not fitting a particular document category (e.g., a document is clearly relevant or not, privileged or not). That seed set of documents is fed into the software to train it for assessing relevancy."); *Biomet I* ("Under predictive coding, the software 'learns' a user's preferences or goals; as it learns, the software identifies with greater accuracy just which items the user wants").

[56] Transcript of Record at 114, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134; Yablon, *supra* note 4, at 639.

[57] Grossman, *supra* note 1, at 29. There are other methods for creating a seed set, the most prominent of which is referred to as "Active Learning" or "Continuous Active Learning." *Id.* at 8 ("An Iterative Training regimen in which the Training Set is repeatedly augmented by additional Documents chosen by the Machine Learning Algorithm, and coded by one or more Subject Matter Expert(s)."); *see also* Grossman, *supra* note 7, at 289-90; Ralph Losey, *Latest Grossman and Cormack Study Proves Folly of Using Random Search For Machine Training – Part Three,* E-DISCOVERY TEAM (July 27, 2014), http://e-discoveryteam.com/2014/07/27/latest-grossman-and-cormack-study-proves-folly-of-using-random-search-for-machine-training-part-three/.

[58] Grossman, *supra* note 1, at 29.

[59] *See* Transcript of Record at 114, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134 (opining that "you should train your algorithm from the kinds of relevant documents that you might actually uncover in a search."); Yablon, *supra* note 4, at 639, 642-43.

[60] *See* Nicholas M. Pace & Laura Zakaras, *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery,* RAND INSTITUTE FOR CIVIL JUSTICE 59 (2012), *available at* http://www.rand.org/pubs/monographs/MG1208.html.

[61] *See* Ralph Losey, *Predictive Coding and The Proportionality Doctrine: A Marriage Made In Big Data,* 26 REGENT U. L. REV. 7, 22 (2013) (describing different sampling methods and the author's preference for using a combined approach).

Random Sampling

A seed set generated by random sampling involves taking a statistically valid sample from the universe of potentially responsive information.[62] That sample, typically created through the functionality of a particular TAR application, is designed to ensure that the seed set reflects the characteristics of the entire document population.[63] Once the randomly comprised seed set is reviewed and coded for responsiveness, the remaining subset of information is run against the universe of documents.[64] That subset may or may not be increased with additional samples to ensure that it is truly representative of the overall percentage of responsive information within the universe of documents.[65] The additional samples could be derived from further rounds of random sampling or from judgmental samples.[66]

Training

Once a seed set has been prepared, it can then be run through the TAR process to train the algorithm and thereby identify responsive information for production.[67] That training process is generally conducted using either of two methods. The first—typically referred to as "simple learning"—trains the algorithm by iteratively running the seed set against the universe of potentially responsive documents.[68] This is done until sufficient information has been obtained to accurately estimate the probability that other unreviewed documents within the document population are relevant to the matter.[69] That subset of unreviewed documents is then reviewed, coded for responsiveness, and (depending on its content) separated for production.

The other training method is generally referred to as "continuous active learning." Like simple learning, continuous active learning technologies rely on a seed set to begin the training process.[70] Where continuous methods differ is that the documents initially identified as responsive are coded and then, depending on the significance of their content, added to the seed set to retrain the algorithm.[71] This process of retraining, identifying new responsive documents, and expanding the seed set with other documents is iteratively repeated until the process does not yield

---

[62] Grossman, *supra* note 1, at 27; Losey, *supra* note 61, at 22; Yablon, *supra* note 4, at 639, 643.

[63] Karl Schieneman & Thomas Gricks, *The Implications of Rule 26(g) on the Use of Technology-Assisted Review*, 7 FED. CTS. L. REV. 239, 260-61 (2013).

[64] Yablon, *supra* note 4, at 639-40.

[65] *Id.* at 643.

[66] Losey, *supra* note 61, at 22.

[67] Transcript of Record at 114, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134; Yablon, *supra* note 4, at 639. The *2016 Guidelines* do not take a position on whether a particular method is better for training a TAR algorithm. Instead, lawyers can use either method to efficiently and effectively generate productions that satisfy standards of reasonableness, and proportionality. *See* Gareth Evans, *Rethinking TAR: New Technologies and Strategies Should Bring the Promise Closer to Reality*, LEGALTECHNEWS (Dec. 15, 2015), *available at* http://www.gibsondunn.com/news/Documents/Evans-Rethinking-TAR-New-Technologies-and-Strategies-Legaltechnews-12-30-2015.pdf.

[68] Simple learning includes "Simple Active Learning" and "Simple Passive Learning." *See* Maura R. Grossman & Gordon V. Cormack, *Continuous Active Learning for TAR*, PRACTICAL LAW THE JOURNAL 32 (April/May 2016) (comparing various aspects of simple learning to "continuous active learning").

[69] *See Biomet I*, at *1 (describing the inextricably intertwined process used for seeding and training the TAR workflow); Remus, *supra* note 2, at 1694, 1702; Yablon, *supra* note 4, at 639.

[70] Grossman, *supra* note 68, at 36-37.

[71] *Id. See generally* John Tredennick et al., *TAR for Smart People*, CATALYST REPOSITORY SYSTEMS (2015) (discussing the functionality of continuous activity learning technologies).

new materially responsive results.[72] The subset of documents that the algorithm identified and that human reviewers coded as responsive are then produced in discovery.

Validation

Finally, counsel should validate the final production results from the TAR process through different forms of testing.[73] This will entail taking statistically valid samples to ensure that the TAR process reached reasonable levels of recall and precision.[74] This result will likely vary depending on the nature of the case, any agreement between the parties, or any court order addressing these issues.[75]

III. SHOULD THE USE OF TAR BE DISCLOSED TO LITIGATION ADVERSARIES?

The issue of whether to disclose the use of TAR is significant and more than just a single inquiry. If counsel is inclined to reveal its use of TAR, how much information will it share? Will counsel merely divulge the fact that TAR will be used, enter into a stipulated use protocol, or adopt an approach that is somewhere between those positions?[76] Beyond the context of traditional litigation, counsel and clients also need to consider the issue of disclosure in connection with government investigations.

Disclosure in Litigation

In traditional litigation, disclosure may be a matter of first impression for many courts,[77] with outcomes influenced by a variety of factors. The history of the judge(s) overseeing the matter,[78] the nature of the client, the temperament of litigation adversaries,[79] and the particular phase of litigation in which TAR is to be used[80] may impact the decision-making process of both counsel and the courts on this issue.

One of the principal factors weighing in favor of some form of disclosure is that of greater certainty, *i.e.*, that an adversarial party will find it difficult to impugn the adequacy of the responding party's search methodology if it is aware that the responding party is using TAR.[81] Against this position stands the counter-argument to disclosure:

---

[72] *Id.*

[73] *See* Byram, *supra* note 28, at 697 ("Sampling furthers the desirability of [TAR] by aiding in confirming a party's results and satisfying the other parties of the contents of the production. In essence, sampling supports defensibility."); *Rio Tinto I*, at 128-29 (discussing some of the methods that requesting parties can use to determine the sufficiency of a production).

[74] Barry, *supra* note 39, at 369-70.

[75] *Id*.

[76] Schieneman, *supra* note 63, at 254-57, 261-63 (describing some considerations surrounding disclosure and transparency with respect to TAR).

[77] *See*, *e.g.*, Transcript of Record at 114, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134 (describing her experience with TAR, U.S. District Judge Denise Cote stated to counsel that "I'm learning about TAR as we go.").

[78] *Id*.; Yablon, *supra* note 4, at 673 (discussing the divergence of views over "how much control courts may and should exercise over discovery practice.").

[79] Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *2 (S.D.N.Y. Aug. 1, 2016) (observing that a responding party decided against entering into a stipulated TAR use protocol given the "history" of its dealings with the requesting party). *See also infra* Part IV.

[80] *See supra* Part II.

[81] *See Dynamo Holdings II,* at 7-9 (approving petitioner's production of documents after respondent raised questions regarding the effectiveness of the parties' cooperatively developed TAR process); Byram, *supra* note 28, at 699 ("Courts will look more favorably upon a party who

potentially costly satellite litigation over an adversary's real or perceived dissatisfaction with the responding party's disclosed use of TAR.[82] That such satellite litigation may occur is evidenced by various cases, the most recent being *Dynamo Holdings v. Commissioner* and *Rio Tinto v. Vale*.[83]

Ultimately, the decision whether to disclose the use of TAR remains in the hands of the responding party and its counsel.[84] This is consistent with the prevailing discovery practice reflected in case law and memorialized in other authoritative sources that the responding party is in the best and most appropriate position to determine how to produce its responsive documents and otherwise satisfy its discovery obligations imposed by the FRCP, local rules, and case law.[85] While it may be strategically beneficial to enter into a cooperative dialogue with the requesting party, doing so may not always be possible or advantageous.[86] Moreover, blanket requirements of transparency are not required by the spirit or letter of the law on this issue.[87]

Disclosure in Government Investigations

Beyond the context of traditional litigation, counsel and clients also need to consider the issue of disclosure in connection with government investigations. This is because government investigators have come to expect that parties will use TAR to facilitate their responsive document productions.[88] As a result, government agencies have in some cases published disclosure requirements or guidelines regarding the use of TAR or other search methodologies.[89]

For example, the Antitrust Division of the U.S. Department of Justice has promulgated specific disclosure guidelines for parties who use TAR "to identify or eliminate potentially responsive documents and information" to respond to a Second Request.[90] Among other things, responding parties are advised to disclose "a detailed description of the

---

discloses its key custodians and how it will [search] for the requested documents. Where a party is transparent, 'opposing counsel and the Court are more apt to agree to your approach . . . .'").

[82] *See infra* Part IV.

[83] *Dynamo Holdings II* (resolving subsequent wrangling by the parties over the effectiveness of the jointly developed TAR process); *Rio Tinto II* (appointing a special master to address the various disputes between the parties arising from their stipulated use protocol regarding TAR).

[84] Kleen Products LLC v. Packaging Corp. of Am., No. 10-cv-5711, 2012 WL 4498465, at *5 (N.D. Ill. Sept. 28, 2012); Ford Motor Co. v. Edgewood Props., 257 F.R.D. 418, 427 (D.N.J. 2009); *SEDONA PRINCIPLES*, at 38.

[85] *Id.*

[86] *See infra* Part IV.

[87] *See, e.g.*, *Biomet II*; Remus, *supra* note 2, at 1716-19; *but see Boardley,* at *11-12.

[88] Robert D. Brown, Jr., *Predictable? — DOJ Approves Use of Predictive Coding in AB InBev-Grupo Modelo Merger Investigation*, GIBBONS E-DISCOVERY LAW ALERT (July 24, 2013), *available at* http://www.ediscoverylawalert.com/2013/07/articles/legal-decisions-court-rules/predictable-doj-approves-use-of-predictive-coding-in-ab-inbev-grupo-modelo-merger-investigation/ (discussing the U.S. Department of Justice's approval of certain parties' use of TAR to produce documents in response to its Second Request).

[89] *See Model Request for Additional Information and Documentary Material (Second Request), Hart-Scott-Rodino Premerger notification Program*, FEDERAL TRADE COMMISSION (Revised August 2015), *available at* https://www.ftc.gov/system/files/attachments/merger-review/guide3.pdf; *Model for Second Request*, U.S. DEPARTMENT OF JUSTICE (June 2015), *available at* https://www.justice.gov/atr/file/706636/download; *Electronic Discovery At The Antitrust Division: An Update*, THE UNITED STATES DEPARTMENT OF JUSTICE (Updated June 25, 2015), *available at* https://www.justice.gov/atr/electronic-discovery-antitrust-division-update; Tracy Greer, *Technology-Assisted Review and Other Discovery Initiatives at the Antitrust Division* (2014), *available at* https://www.justice.gov/sites/default/files/atr/legacy/2014/03/27/304722.pdf.

[90] *Model for Second Request*, U.S. DEPARTMENT OF JUSTICE, at *13 (June 2015), *available at* https://www.justice.gov/atr/file/706636/download.

method(s) used to conduct all or any part of the search."[91] This could include the "structure" of their TAR "workflow," the methods used to develop a "seed set," and validation testing results.[92]

In addition, responding parties may need to provide "a statistically significant sample of both relevant and non-relevant documents" and devise a methodology to ensure the production of "additional relevant information" if required by "an issue . . . that had not been anticipated by the Division at the outset."[93] Explicit directives of this nature suggest that parties who wish to use TAR to respond to an Antitrust Division inquiry should carefully follow the published guidelines to avoid protracting or otherwise complicating an investigation.[94]

The Federal Trade Commission (FTC) has also provided direction on the use of TAR by parties who are tasked with responding to Second Requests from the FTC.[95] In its Model Second Request, the FTC seeks a comprehensive disclosure of the responding party's TAR process. This includes key aspects and metrics associated with the "collection methodology" such as:

- The manner in which the technology was used to "to identify responsive documents"
- The seed set development process
- The number of documents that were manually reviewed
- Those documents deemed nonresponsive "without manual review"
- The validation process for responsive and nonresponsive documents and
- How the responding addressed any "foreign language documents."[96]

Beyond all of this, the responding party should be prepared to disclose "all statistical analyses" relating to "the precision, recall, accuracy, validation, or quality" of the documents produced in response the Second Request.[97]

The guidance from the DOJ and FTC ultimately teaches that counsel should use TAR in such a way that will both protect and advance the client's interests in an investigation. With government agencies specifically admonishing responding parties "against using TAR without disclosing its use," that may very well mean full disclosure of the entire TAR process.[98] For other government investigations, it may or may not require full disclosure. Decisions on this issue should be made only after carefully studying agency requirements and practices, including the willingness of government agencies to modify production requirements based on the specific circumstances of an

---

[91] *Id*.

[92] Tracy Greer, *Technology-Assisted Review and Other Discovery Initiatives at the Antitrust Division*, at *3-4, *available at* https://www.justice.gov/sites/default/files/atr/legacy/2014/03/27/304722.pdf.

[93] *Electronic Discovery At The Antitrust Division: An Update*, THE UNITED STATES DEPARTMENT OF JUSTICE, at *4-5 (Updated June 25, 2015), *available at* https://www.justice.gov/atr/electronic-discovery-antitrust-division-update.

[94] *Id*.

[95] *See Model Request for Additional Information and Documentary Material (Second Request), Hart-Scott-Rodino Premerger notification Program*, FEDERAL TRADE COMMISSION, at *15-16 (Revised August 2015), *available at* https://www.ftc.gov/system/files/attachments/merger-review/guide3.pdf. While the FTC's announcement publishing its Model Second Request references "predictive coding," the Model Second Request instead uses the term "technology-assisted review" to encompass predictive coding technology. *Id*. at *19 ("The term 'Technology Assisted Review' means any process that utilizes a computer algorithm to limit the number of potentially responsive documents subject to a manual review.").

[96] *Id*. at *16.

[97] *Id*.

[98] Tracy Greer, *Technology-Assisted Review and Other Discovery Initiatives at the Antitrust Division*, at *6, *available at* https://www.justice.gov/sites/default/files/atr/legacy/2014/03/27/304722.pdf.

investigation.[99] Client needs and interests, along with other related factors, should also be taken into consideration.[100]

IV. WHAT ARE THE BENEFITS AND DRAWBACKS OF ENTERING INTO A STIPULATED TAR USE PROTOCOL?

A critical decision for counsel is whether to enter into a stipulated protocol regarding the use of TAR. Counsel's choice on this issue will affect the course of discovery, impact the relationship between the parties, and influence the court's perception of counsel and client.[101]

It is worth emphasizing that it is the choice of counsel and the client to enter into such a protocol.[102] While some courts and commentators have taken the position that parties should enter into a stipulated protocol to use TAR,[103] neither the Rules nor case law require such a step[104] unless ordered by the court.[105] Indeed, whether counsel should take this step is entirely dependent on what is best for its client,[106] not what is convenient for opposing counsel or the court, so long as the production satisfies the prevailing standards of relevance, proportionality, and reasonableness.[107] While there are potential benefits to entering into such a protocol, there are also risks. The purpose of this section is to outline issues regarding the use of stipulations so counsel can make an informed decision that best represents client interests.

---

[99] *See Model Request for Additional Information and Documentary Material (Second Request), Hart-Scott-Rodino Premerger notification Program*, FEDERAL TRADE COMMISSION, *4 (Revised August 2015), *available at* https://www.ftc.gov/system/files/attachments/merger-review/guide3.pdf (If the Company believes that the required search or any other part of the Request can be narrowed in any way that is consistent with the Commission's need for documents and information, you are encouraged to discuss any questions and possible modifications with the Commission representatives).

[100] When deciding how to approach the use of TAR in connection with a government investigation, counsel and clients would be well advised to keep in mind the famous admonition from Oliver Wendell Holmes Jr. that "[m]en must turn square corners when they deal with the government." Rock Island, Ark. & La. RR v. United States, 254 U.S. 141, 143 (1920).

[101] *See, e.g.*, Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 204 (S.D.N.Y. 2012), In re Actos (Pioglitazone) Products Liab. Litig., 6:11-MD-2299, 2012 WL 7861249 (W.D. La. July 27, 2012); Transcript of Record at 110-111, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134.

[102] *See supra* Part III.

[103] *See, e.g., Da Silva Moore*, 287 F.R.D. at 192; Byram, *supra* note 28, at 699.

[104] FED. R. CIV. P. 26(b); 34(b); Kleen Products LLC v. Packaging Corp. of Am., No. 10-cv-5711, 2012 WL 4498465, at *5 (N.D. Ill. Sept. 28, 2012); Ford Motor Co. v. Edgewood Props., 257 F.R.D. 418, 427 (D.N.J. 2009); *SEDONA PRINCIPLES*, at 38.

[105] Order Re: Implementation of Predictive Coding Regimen, Indep. Living Ctr. of S. Cal. v. City of L.A., 2:12-cv-00551-FMO-PJW, at *1 (C.D. Cal. June 13, 2014) ECF No. 371 (ordering the defendant "to use a predictive coding system for identifying the 10,000 most relevant documents in its databases and, after reviewing them for privilege, etc., produc[e] them to Plaintiffs); Memorandum and Order, Aurora Co-op. Elevator Co. v. Aventine Renewable Energy – Aurora West, LLC, 4:12-cv-230 (D. Neb. Mar. 10, 2014) ECF No. 147 (ordering the parties to use "predictive coding" in connection with the first phase of discovery and, in connection with that process, to designate an expert who could work with the parties to design a computerized search to examine the parties' electronic records).

[106] MODEL RULES OF PROF'L CONDUCT PREAMBLE & SCOPE ¶ 2 (2013) (explaining generally a lawyer's duties to its client including, but not limited to, the requirement to "zealously assert[] the client's position under the rules of the adversary system").

[107] *See supra* Parts I, II.

The Principal Benefit of a Stipulated Use Protocol is Cost Savings

The primary objective of entering into a stipulated TAR protocol is generally to decrease the costs associated with pursuing discovery.[108] Proponents of this strategy argue that costs can be reduced since discovery will—in theory—proceed in a more orderly fashion with the court and all parties cooperatively involved in the process.[109] Many stipulated protocols have invited opposing counsel to collaborate with and help prepare its adversary's search methodology.[110] This may include sharing irrelevant documents with opposing counsel from the sample and seed sets or allowing counsel to assist with document coding. Opposing counsel may also be invited to participate in the training and testing processes.[111] According to its proponents, such a cooperative and transparent approach will reduce satellite litigation over the process the party used to search for, review, and produce responsive information.[112] All of which will arguably make discovery less costly, more efficient, and ultimately focused on disclosing information to enable the parties to resolve matters on the merits.[113]

Understanding the Drawbacks of a Stipulated Protocol

Against the backdrop of potentially lower discovery costs loom several drawbacks with stipulated use protocols. The first and most obvious risk is the potential for excessive input from and wrangling with opposing counsel and the court over the process for searching, reviewing, and producing documents.[114] The ESI search and review process has always been complex; allowing opposing counsel to participate may create tensions given the parties' adversarial interests in the litigation.[115] Along with those tensions is the likelihood of motion practice and delays, which can offset the cost and time savings otherwise offered by TAR.[116]

Another hazard with stipulated use protocols is that they may require counsel to disclose to its litigation adversary non-responsive information, particularly non-responsive documents used to train the TAR algorithm.[117] Besides the fact that such information is outside the permissible scope of discovery,[118] disclosing non-responsive documents

---

[108] *See*, *e.g.*, L. Casey Auttonberry, Comment, *Predictive Coding: Taking the Devil Out of the Details*, 74 LA. L. REV. 613, 624 (2014) ("Well-designed protocols can effectively decrease some of the costs and delays associated with e-discovery.").

[109] *See generally Da Silva Moore*, 287 F.R.D. 182; *In Re Actos,* 2012 WL 7861249.

[110] *Id. See also* Judge Henry Coke Morgan, Jr., *Predictive Coding: A Trial Court Judge's Perspective*, 26 REGENT U.L. REV. 71, 77-78 (2013) (opining that "counsel for all parties will participate in the two-step process of selecting the seed set of documents for coding . . . . [which] must be transparent and acceptable to all parties."). *Contra* Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *2 (S.D.N.Y. Aug. 1, 2016) ("The City declined, both because of cost and concerns that the parties, based on their history of scope negotiations, would not be able to collaborate to develop the seed set for a TAR process.").

[111] *Da Silva Moore*, 287 F.R.D. 182; *In Re Actos,* 2012 WL 7861249. *But see Biomet II*, at *1 ("The Steering Committee wants the whole seed set Biomet used for the algorithm's initial training. That request reaches well beyond the scope of any permissible discovery by seeking irrelevant or privileged documents used to tell the algorithm what not to find. That the Steering Committee has no right to discover irrelevant or privileged documents seems self-evident.").

[112] Schieneman, *supra* note 63, at 261-63; Byram, *supra* note 28, at 698-699.

[113] *Id*.

[114] Transcript of Record at 110-111, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134; *Da Silva Moore*, 287 F.R.D. at 185-189.

[115] *Id.*; *Hyles*, 2016 WL 4077114, *2. *But see Sedona Conference Cooperation Proclamation*, 10 SEDONA CONF. J. 331 (Fall Supp. 2009).

[116] This is particularly evident from the docket in *Da Silva Moore*, which reflects repeated motion practice over nominal issues relating to the parties' TAR protocol. *See* Facciola, *supra* note 51, at 18, n. 141.

[117] *See* Remus, *supra* note 2, at 1716-19.

[118] FED. R. CIV. P. 26(b)(1); *Biomet II*, at *1-2.

may violate counsel's ethical duty of confidentiality.[119] Non-responsive information may contain trade secrets, sensitive financial data, or other proprietary information that should not be disclosed to a litigation adversary,[120] especially if the adversary is a business competitor.[121] Moreover, depending on the nature of the information, it could be used to add new claims in the present lawsuit[122] or to file a new lawsuit against the producing party.[123]

A third problem associated with stipulated protocols is the risk of waiving attorney work product protection by voluntarily identifying the TAR seed set to opposing counsel.[124] A seed set is a selection of documents that may reflect a lawyer's perceptions of relevance, litigation tactics, or even its trial strategy.[125] These selections of documents have been protected as fact or even opinion work product.[126] By pinpointing those specific documents for a litigation adversary, counsel likely yields any work product protection that may have otherwise been associated with its identification of those documents.[127]

A related risk with stipulated protocols is the possibility that attorney-client privileged information could be inadvertently shared with opposing counsel.[128] This is particularly the case where privileged communications are used to train the TAR process.[129] Unless appropriate screening measures are deployed, counsel could inadvertently

---

[119] *See*, *e.g.*, MODEL RULES OF PROF'L CONDUCT R. 1.6 (2013) (delineating the general rule that "[a] lawyer shall not reveal information relating to the representation of a client," along with pertinent exceptions); CAL. BUS. AND PROF. CODE § 6068(e) ("It is the duty of an attorney to . . . maintain inviolate the confidence, and at every peril to himself or herself to preserve the secrets, of his or her client."); STATE BAR OF CALIFORNIA STANDING COMM. ON PROF'L RESPONSIBILITY AND CONDUCT, FORMAL OP. NO. 2015-193 (2015), *available at* https://ethics.calbar.ca.gov/Portals/9/documents/Opinions/CAL%202015-193%20%5B11-0004%5D%20(06-30-15)%20-%20FINAL.pdf

[120] *Id*.

[121] *Cf.* Apple v. Samsung, 5:11-cv-01846, 2014 WL 2854994, at *4 (N.D. Cal. June 20, 2014) (imposing a $2 million sanction on the defendant's counsel (among others) for making an unauthorized disclosure to its client of documents produced by the plaintiff that were designated "attorneys' eyes only" under the governing protective order).

[122] *Cf.* In re Google Inc., 462 F. App'x 975, 978 (Fed. Cir. 2012) (the defendant's unintended disclosure of certain documents claimed as privileged arguably strengthened the plaintiff's claims).

[123] *Cf.* Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc., 2014 WL 584300, at *3 (S.D.N.Y. Feb. 14, 2014) (rejecting certain defendants' attempt to use documents produced in a related action to challenge the reasonableness of the plaintiff's document production in the instant action).

[124] *See Biomet II*; Remus, *supra* note 2, at 1716-17 ("requiring seed-set transparency threatens core protections for attorney work product, attorney–client privilege, and confidentiality."); Yablon, *supra* note 4, at 644 ("If . . . the seed set is made up of documents selected or coded by a producing party as relevant, production of that seed set has a much higher probability of disclosing attorney impressions of the case.").

[125] *See generally* Facciola, *supra* note 51 (detailing the circumstances under which seed sets may be protected as attorney work product).

[126] FED. R. CIV. P. 26(b)(3)(B) ("If the court orders discovery of those materials, it must protect against disclosure of the mental impressions, conclusions, opinions, or legal theories of a party's attorney or other representative concerning the litigation."); Hickman v. Taylor, 329 U.S. 495, 516 (1947) (Jackson, J., concurring) ("Discovery was hardly intended to enable a learned profession to perform its functions either without wits or on wits borrowed from the adversary"); Lockheed Martin Corp. v. L-3 Communs. Corp., 6:05-cv-1580-Orl-31KRS, 2007 WL 2209250, at *9 (M.D. Fla. July 29, 2007) (explaining that opinion work product – which includes "an attorney's mental impressions, conclusions, opinions, or legal theories" – is entitled to "a nearly absolute immunity and can be discovered only in very rare and extraordinary circumstances").

[127] Simmons, Inc. v. Bombardier, Inc., 221 F.R.D. 4, 8 (D.D.C. 2004) ("The work-product privilege may be waived by the voluntary release of materials otherwise protected by it.").

[128] *See*, *e.g.*, Transcript of Record at 14-15, Fed. Hous. Fin. Agency v. JPMorgan Chase & Co., Inc., 11-cv-05201 (S.D.N.Y. July 24, 2012) ECF No. 128.

[129] *Id*. (addressing the issue of privilege protection in connection with sharing seed set documents with the plaintiff's counsel, counsel for one of the defendants requested the court "to make some provision, your Honor, which I think would be necessary, to deal with the problem that privileged documents could be part of the seed set, and when you don't have plaintiff's counsel in the picture, you don't have to worry about maintaining the privilege, but once you do have them in the picture, you'd have to take a step or two to make sure that that wasn't going to be an issue.").

divulge privileged materials to opposing counsel.[130] And while properly executed orders under Federal Rule of Evidence 502(d) may address the problem of inadvertent waiver,[131] they cannot remove the nature of the privileged information now in the mind of opposing counsel.[132]

Finally, any deviations from the stipulated protocol that counsel may want to implement will likely require consent from opposing counsel or the court.[133] Implementing unilateral changes to the agreed-upon process may invite the court's displeasure and adverse consequences to counsel and the client.[134]

The Jurisprudence on Stipulated Protocols

There are few reported cases that address the benefits and risks of entering into a stipulated TAR protocol. In particular, the extant jurisprudence illustrates the risks of those protocols. This is not surprising since "reported decisions tend to involve obstructionist conduct at the most egregious end of the spectrum."[135] The following three cases are instructive on these issues.

Rio Tinto v. Vale

In *Rio Tinto*, the court entered an order approving the parties' stipulation to use TAR.[136] That stipulation—referred to by the parties as their "TAR Protocol"—was generally designed to make the use of TAR in *Rio Tinto* self-executing.[137] In other words, the parties should not need constant supervision from and intervention by the court. However, the court's docket reveals that the parties subsequently sought judicial relief on a variety of disputes arising from the Protocol.

For example, the court resolved the parties' dispute over the use of search terms to cull down the document universe before using TAR.[138] While the court permitted defendant Vale to use search terms consistent with the TAR

---

[130] *Id*.

[131] FED. R. EVID. 502(d); *see also* FED. R. EVID. 502 advisory committee's note (discussing the rule's framework for addressing the problems associated with the inadvertent production of ESI); John M. Barkett, *Evidence Rule 502: The Solution to the Privilege-Protection Puzzle in the Digital Era*, 81 FORDHAM L. REV. 1589, 1619-20 (2013) (discussing the importance of Federal Rule of Evidence 502(d) in reducing the costs and burdens associated with attorney-client privilege reviews in discovery).

[132] Philip Favro, *Perspective: Oracle v. Google Trial Highlights Importance of Privilege Reviews,* BLOOMBERG (May 11, 2016), https://bol.bna.com/perspective-oracle-vs-google-trial-highlights-importance-of-privilege-reviews/.

[133] *Compare Boardley* at *11-12 ("Parties . . . are expected to cooperate in the . . . use of computer-assisted search methodology, such as Technology Assisted Review . . . The failure of a party or counsel to cooperate will be relevant in resolving any discovery disputes . . . [and] in determining whether the Court should impose sanctions in resolving discovery motions") *with* Connecticut General Life Ins. Co. v. Health Diagnostic Laboratory, Inc., No. 3:14-cv-01519-VAB, *4 (D. Conn. Jan. 28, 2015) ("No party is required to use search terms; further, use of search terms does not foreclose a party's right to use other technologies, where appropriate. It is the producing party's right to determine the best method of search and review and how it will comply with Rule 26.").

[134] *Boardley*, at *11-12. *See also Biomet II* (warning that "[a]n unexplained lack of cooperation in discovery can lead a court to question why the uncooperative party is hiding something. . . .").

[135] Robert Douglas Brownstone, *Collaborative Navigation of the Stormy e-Discovery Seas*, 10 RICH. J.L. & TECH. 53, ¶ 29 (2004).

[136] *Rio Tinto I*, at 129.

[137] *Id*. at 129-131.

[138] *See* Joint Letter, Rio Tinto v. Vale, 1:14-cv-03042, at *7 (S.D.N.Y. May 6, 2015) ECF No. 246 (reflecting a joint letter by the parties to the court that memorializes the court's resolution of the disputed issue); *See* Joint Letter, Rio Tinto v. Vale, 1:14-cv-03042, at *5-6, *16-17 (S.D.N.Y. Apr. 6, 2015) ECF No. 234 (containing a joint letter by the parties to the court that reflects the parties' dispute and their respective arguments).

Protocol, it nonetheless allowed plaintiff Rio Tinto to propose additional search terms that might capture relevant information that could otherwise have been excluded.[139]

While the court managed to address this dispute, the parties continued to quarrel over other process and training issues.[140] The need to resolve those matters and possibly other, future TAR disputes apparently led the court to appoint a special master.[141] While that step yielded a temporary truce in the form of *another* stipulation and order regarding the use of TAR,[142] it came only after an unusually explicit judicial admonishment that the parties "learn to follow Fed. R. Civ. P. 1 . . . [and] to cooperate more."[143]

That a special master would be required to help the parties—particularly where they voluntarily entered into a use protocol—seems counterintuitive and underscores the need for counsel to proceed with caution before agreeing to such protocols. Indeed, *Rio Tinto* teaches that parties should carefully assess whether and what to disclose regarding their use of TAR.[144] While it may be strategically beneficial to enter into a cooperative dialogue with the requesting party, doing so may not always be advantageous.[145] Before taking any action on this front, counsel should first understand the background of the judge overseeing the matter,[146] the nature of the client, the temperament of its litigation adversary, the cooperativeness of opposing counsel,[147] and the particular phase of litigation in which TAR is to be used. Such a course will likely help the client make an informed and effective decision on the issues.

Progressive Casualty Insurance Co. v. Delaney

In *Progressive*, the court took the unusual step of prohibiting the plaintiff from using TAR to search through and review its documents.[148] The parties had initially entered into a keyword search protocol that was reflected in the court's case management order.[149] Pursuant to that order, the plaintiff agreed to run a series of keyword searches jointly developed with the defendants against its universe of potentially responsive information.[150] Once it

---

[139] *See* Joint Letter, Rio Tinto v. Vale, 1:14-cv-03042, at *7 (S.D.N.Y. May 6, 2015) ECF No. 246.

[140] *Rio Tinto II*, at *1 ("There still are issues regarding the parties' TAR-based productions (including an unresolved issue raised at the most recent conference)").

[141] *Id*. at *1-2.

[142] Stipulation and Order Re: Revised Validation and Audit Protocols for The Use of Predictive Coding, Rio Tinto v. Vale, 1:14-cv-03042 (S.D.N.Y. Sep. 8, 2015) ECF No. 338.

[143] Memo Endorsed Order, Rio Tinto v. Vale, 1:14-cv-03042, (S.D.N.Y. Aug 13, 2015) ECF No. 319.

[144] *See supra* Part III.

[145] *See id*.; MODEL RULES OF PROF'L CONDUCT PREAMBLE & SCOPE ¶ 2 (2013).

[146] *See Da Silva Moore*, 287 F.R.D. at 192 ("[T]ransparency allows the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called 'black box' of the technology. This Court highly recommends that counsel in future cases be willing to at least discuss, if not agree to, such transparency in the computer-assisted review process.").

[147] *See, e.g.*, Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *2 (S.D.N.Y. Aug. 1, 2016); Transcript of Record at 110-111, Fed. Hous. Fin. Agency v. UBS Americas, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134. Counsel for defendant JPMorgan Chase & Co. summarized the difficulties of working jointly with the plaintiff in developing a TAR workflow: "We meet every day with the plaintiff to have a status report, get input, and do the best we can to integrate that input. It isn't always easy, not just to carry out those functions but to work with the plaintiff. The suggestions we have had so far have been unworkable and by and large would have swamped the project from the outset and each day that a new suggestion gets made. But we do our best to explain that and keep moving forward." *Id*.

[148] Progressive Cas. Ins. Co. v. Delaney, No. 2:11-cv-00678, 2014 WL 3563467, at *10-12 (D. Nev. July 18, 2014).

[149] *Id.* at *6-9.

[150] *Id.*

completed those searches, the plaintiff would either manually review the subset of potentially responsive documents or simply produce all of the documents except for those that were potentially privileged.[151] While this protocol seemed acceptable at first, the plaintiff quickly determined that it would be too expensive to manually review the over 500,000 documents that remained after running the keyword searches.[152]

The plaintiff then turned to TAR given its potential to expedite the search process at a reduced cost.[153] However, the plaintiff did not initially disclose to opposing counsel its decision to modify its keyword search protocol and neglected to produce any responsive documents in the meantime.[154] When the parties subsequently failed in their attempt to reach an agreement on the use of TAR, the court placed the blame on the plaintiff.[155] Not only did the court forbid the plaintiff from using TAR, it ordered the plaintiff to make a blanket production of the potentially responsive documents within two weeks, with the exception that it could withhold arguably privileged materials.[156]

The fundamental lesson from *Progressive* concerns unilaterally violating an agreement entered as an order of the court.[157] While the *Progressive* case dealt with a keyword search protocol, its holding is equally applicable to TAR use stipulations. Once such a stipulation is reached and reduced to a court order, it may be difficult to change.[158] If counsel becomes dissatisfied with the framework governing the TAR process, *e.g.*, the level of its adversary's involvement in the training of the process, it may be foreclosed from unilaterally modifying its strategy and tactics just like the plaintiff from *Progressive*.[159] Instead, counsel would likely have to negotiate with its adversary or ask the court to modify the agreement.[160] All of which could be costly to the client and could still fail to yield the sought-after results.

Like *Rio Tinto*, *Progressive* spotlights the importance of carefully examining the pertinent circumstances surrounding a particular matter before deciding that a stipulated use protocol is the best course of action.

Connecticut General Life Insurance Company v. Health Diagnostic Laboratory, Inc.

Connecticut General involved a scenario where the parties entered into a stipulation to generally address the collection and production of ESI.[161] Unlike *Rio Tinto*, the stipulation in *Connecticut General* was not a TAR use

---

[151] *Id.*

[152] *Id.* at *2-5.

[153] *Id.*

[154] *Id.* at *9.

[155] *Id.* at *10-12.

[156] *Id.*

[157] *Id.*

[158] *Id.*

[159] *Contra Bridgestone*, at *1-2.

[160] *See* IBM's Response to Bridgestone's Request to Add Predictive Coding to the ESI Protocol, Bridgestone Americas, Inc. v. Int'l Bus. Mach. Corp., No. 3:13-cv-1196 (M.D. Tenn. July 14, 2014) ECF No. 85; Bridgestone's Reply to IBM's Opposition to Use of Attorney Trained Technology-Assisted Review, Bridgestone Americas, Inc. v. Int'l Bus. Mach. Corp., No. 3:13-cv-1196 (M.D. Tenn. July 18, 2014) ECF No. 88; Progressive Cas. Ins. Co. v. Delaney*, No. 2:11-cv-00678, 2014 WL 3563467, at *4, *8-11 (D. Nev. July 18, 2014).

[161] Connecticut Gen. Life Ins. Co. v. Health Diagnostic Laboratory, Inc., No. 3:14-cv-01519, 2015 WL 471720 (D. Conn. Jan. 28, 2015).

protocol. Instead, the stipulation tackled search issues more broadly and adopted a more flexible approach to search methodologies than the stipulation that proved so troublesome in *Progressive*.

In particular, the *Connecticut General* stipulation did not mandate that the parties use a particular search method. Instead, the stipulation provided that the responding party had the "right to determine the best method of search and review and how it will comply with Rule 26."[162] If a party sought to use keywords, the stipulation provided a specific meet and confer framework for the parties to reach agreement on "the search terms to be used and the sources of ESI to be searched."[163]

Significantly, a party's use of search terms would not foreclose its right to use other search methodologies: "No party is required to use search terms; further, use of search terms does not foreclose a party's right to use other technologies, where appropriate."[164] If the responding party chose "to search and review using a technology or methodology other than search terms (including, for instance, TAR)," it need only divulge "its intent to use that technology and the name of the review tool."[165] No further disclosure obligations would be imposed unless the requesting party established a fact-specific showing of good cause that the responding party's production was somehow deficient.[166]

Unlike *Progressive*, the parties' approach to the discovery search process in *Connecticut General* has not devolved into protracted motion practice. While the success of the parties' stipulation in *Connecticut General* is still to be determined,[167] their approach to search methodologies is instructive since it demonstrates that lawyers can adopt a transparent and cooperative approach to the search and review process that may be beneficial to both parties. For the responding party, such an approach preserves its right to determine how it should address its discovery obligations while avoiding many of the risks of disclosure. It also decreases the chances that a requesting party could successfully object to the use of TAR on the basis of prior non-disclosure.

Nevertheless, the requesting party benefits from such an arrangement since it has learned of the use of TAR and the technology the responding party will use. Having such information should enable the requesting party to better engage with the responding party regarding the completeness of its production and devise methods to ensure that the production process was both reasonable and proportional under the circumstances.

Whether the *Connecticut General* approach is appropriate in a given matter will ultimately depend on that same multiplicity of factors discussed in connection with the *Rio Tinto* case. To ensure that a client's interests are properly represented in discovery, counsel should carefully investigate at the outset of litigation which search and review protocols will best address the demands of discovery in the case.[168] Counsel should also gauge the cooperativeness of opposing counsel and determine the level of involvement that its adversary should have in the search and review

---

[162] *Id.* at 4.

[163] *Id.*

[164] *Id.*

[165] *Id.*

[166] *Id.*

[167] The matter has been stayed due to the defendant's filing of a chapter 11 bankruptcy petition. Order Staying Case, Connecticut Gen. Life Ins. Co. v. Health Diagnostic Laboratory, Inc., No. 3:14-cv-01519 (D. Conn. June 11, 2015) ECF No. 75.

[168] *See* Byram, *supra* note 28, at 697.

process.[169] Finally, counsel should ensure that its client's interests, including its secrets and confidences, are protected, particularly if it decides to enter into a stipulated use protocol.[170]

By addressing these issues in a prudent fashion, counsel will better ensure that it avoids the risks exemplified in the *Rio Tinto* and *Progressive* cases.

V. CONCLUSION

The complexities and confusion surrounding the use of TAR need not obscure the point that TAR is a viable, judicially approved, and dynamic method for conducting discovery. So long as its workflow meets the Rule 1 mandate and the related notions of relevance, proportionality, and reasonableness, use of TAR should be defensible. To adequately satisfy these standards does not always require counsel to enter into a stipulated use protocol or to disclose its use of TAR. Those decisions may or may not be advisable depending on the circumstances of a case.

To establish defensibility, counsel must accurately determine the prevalence of responsive information, ensure that its training process yields acceptable levels of recall and precision for its production of documents, and validate its final production results. By addressing these issues and by deploying safeguards to protect against disclosures of non-responsive information, attorney work product, or attorney-client privileged materials, counsel should have the elements in place to successfully use TAR in discovery.

---

[169] Hyles v. New York City, 10-cv-3119, 2016 WL 4077114, *2 (S.D.N.Y. Aug. 1, 2016).

[170] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6 (2013); STATE BAR OF CALIFORNIA STANDING COMM. ON PROF'L RESPONSIBILITY AND CONDUCT, FORMAL OP. NO. 2015-193 (2015), *available at* https://ethics.calbar.ca.gov/Portals/9/documents/Opinions/CAL%202015-193%20%5B11-0004%5D%20(06-30-15)%20-%20FINAL.pdf.